Retail Investor Attention and Stock Liquidity

Retail Investor Attention and Stock Liquidity*

Rong Ding and Wenxuan Hou†

Abstract

We use the search volume index (SVI) of the stock ticker provided by Google Trends to capture the active attention that retail investors pay to stocks. Based on the analysis of S&P 500 stocks from 2004 to 2009, we show that the majority of the variation in SVI cannot be explained by passive attention measures, including Google News coverage and advertising expenditure. We find that retail investor attention, reflected by the level and change in SVI, significantly enlarges the shareholder base and improves stock liquidity. The results are robust to the control of endogeneity issues.

***Keywords*:** Investors' attention; breadth of ownership; liquidity; bid-ask spread; SVI; media coverage; Google; retail investor

JEL Classification: G10, G14

Journal of International Financial Markets, Institutions & Money, forthcoming

* The authors are grateful for valuable comments from an anonymous reviewer, John Doukas, Angelica Gonzalez, Michela Verardo and the participants in the EFA (European Finance Association) 2011 meeting in Stockholm for helpful comments. We are grateful to Lei Zhang for his assistance with the collection of the SVI data.

† Rong is from the Warwick Business School, University of Warwick, UK; and Wenxuan is from the University of Edinburgh Business School, UK.
Address correspondence to Wenxuan Hou, 29 Buccleuch Place, Edinburgh, EH8 9JS, UK. Tel: +44 (0) 131 651 5319, Fax: +44 (0) 131 651 4389; Email: wenxuan.hou@ed.ac.uk

Highlights

- Search volume index (SVI) of the stock ticker provided by Google Trends captures the active attention of retail investors
- Retail investor attention enlarges the shareholder base
- Retail investor attention improves stock liquidity.

1. Introduction

The "under-diversification puzzle" documented in the literature shows that investors have "home bias" because they tend to favor investment in firms they are familiar with (French and Poterba, 1991; Tesar and Werner, 1995; Cao, Han, Hirshleifer and Zhang, 2011). In order to get familiar with such firms, investors have to spend time and effort collecting relevant information, which suggests that attention from investors might predict the subsequent trading activity. On the theoretical side, studies on asset pricing posit that investor attention is a necessary condition for a stock price to fully reflect public information, as investors need to be aware of the information before they can analyze and react to it (Hirshleifer and Teoh, 2003; Hou, Peng and Xiong, 2008; Hirshleifer, Lim and Teoh, 2011). However, because of the limits on the information-processing capacity of human beings, attention is largely concentrated on the stocks that investors are interested in or familiar with, which implies that attention paid to stocks by investors could result in subsequent trading of these stocks. Our study aims to provide fresh insights into the capital market consequences of investors' attention.

Building on the assumption that the investors passively attend to publicly available information, previous studies have used advertising expenditure (Grullon, Kanatas and Weston, 2004) and media coverage (Fang and Peress, 2009) to capture investors' attention and examine its implications for stock liquidity and stock returns. In this paper, we employ a measure of active attention from investors, recently developed by Da, Engelberg and Gao (2011), namely the aggregate search volume index (SVI) provided by Google Trends (available from: www.google.com/trends), and test the impact of investors' attention paid to listed firms on two aspects of listed firms: breadth of ownership and liquidity. After

controlling for the passive attention measures documented in the literature, we find that increased investors' attention measured by the SVI contributes to a broader shareholder base. This is in line with the argument of Barber and Odean (2008) that retail investors tend to search for information about the firm's history, product, environment and strategies when selecting stocks, and can be interpreted with the "investor recognition hypothesis" (Merton, 1987), which states that the shareholder base measures the recognition of the firm among investors, so that an enlarged shareholder base indicates that the firm has been well recognized. In other words, potential investors have to be aware of a firm before they can gradually become familiar with it and then eventually decide to invest, suggesting that investor attention is a necessary condition for a firm to be recognized. The impact of passive attention measures, however, is not always significant in the results, showing that retail investors do not necessarily invest in firms with more advertising expenditure or media coverage.[3] Furthermore, we find that increased investors' attention, as measured by the SVI, results in reduced bid-ask spread, and our results remain consistent after controlling for the passive attention measures, firm characteristics, and year and industry fixed effects. Our findings remain robust to alternative liquidity measures, including effective spread, relative effective spread, and turnover rate (trading volume divided by shares outstanding).

This paper makes three important contributions to the literature. First, our study contributes to the broad literature on the "investor recognition hypothesis" (i.e., Merton, 1987, Grullon et al., 2004; Tetlock, 2010; Fang and Peress, 2009).[4] Merton (1987) asserts that "ceteris paribus,

[3] We argue that the SVI captures investor attention in a more timely and accurate manner than passive attention measures for the following reasons: 1) media coverage of a firm is sporadic, while the SVI is continuous; 2) media coverage does not necessarily guarantee attention unless investors attend to it, and the same news coverage could generate different levels of investor attention for different stocks (Da et al., 2011).

[4] Empirical evidences largely support the investor's recognition hypothesis. Chen, Noronha and Singal (2004) report an increased investor's awareness after a firm is added to the S&P 500 index, which leads to a reduction in both the information asymmetry component of the bid-ask spread and the Merton (1987)'s cost of under-diversification. By the same token, Lehavy and Sloan (2008) contend that an exchange listing increases

an increase in the relative size of the firm's investor base will reduce the firm's cost of capital and increase the market value of the firm." A stock's visibility is associated with its price, publicity and popularity of the core products and social image. However, we suggest that these measures are passive, in that it is implicitly assumed that firms with high visibility will attract more attention from investors which is difficult to empirically verify. Our study is built on an active measure of *ex post* attention, as Google search is a confirmed measure of attention: if an individual intentionally searches for information about a stock, it is evident that one is paying attention to it (Da et al., 2011).[5] Furthermore, Google search index captures investors' attention in a more timely way than passive measures of attention. When individual investors actively search for a stock using Google, they acquire useful information relevant to the stock, which mitigates the information asymmetry problem for these stocks. As a result, liquidity improves for stocks with better investor recognition.

Second, our paper adds to the emerging literature on investor attention and asset pricing dynamics, including Barber and Odean (2008) on investor attention and individual investors' trading behavior, Engelberg and Parsons (2011) on the casual impact of local media coverage on local trading, Da et al. (2011) on the impact of active attention on IPO returns and price changes in subsequent periods and Aouadi, Arouri and Teulon (2013) on the effect of investor attention on stock market liquidity and volatility use Google French data.

---

investor's recognition of a firm. Furthermore, a positive association between investor's recognition and contemporaneous stock return id documented. Bushee and Miller (2012) find that small and mid-cap firms can enhance their visibility among investors and analysts by hiring an investor relation firm, which contributes to improved market valuation.

[5] Our study is related to, but different from, Grullon et al. (2004) because our paper focuses on the relation between investors' active attention (to a stock) and the firm's shareholder base as well as its liquidity, while Grullon et al. (2004) investigate firms' advertising expenditure, as a (passive) approach used to reach a broad audience, and its impact on breadth of ownership and liquidity.

Finally, our study extends the literature on the stock market consequence of investors' information demand. For example, Vlastakis and Markellos (2011) use the Google search volume of constituents of Dow Jones Industrial Average Index as a proxy of investors' information demand, and find that such information demand has significant impact on stock trading volume and the conditional variance of excess return. Siganos (2013) use Google search volume of target firms involved in a merger between 2004 and 2010 in the UK as a proxy for investor's information demand for the target firms, and find that such measure can explain a large percentage of the price increase in target firms prior to the merger. Vozlyublennaia (2014) use Google search to proxy for investor attention (investors' information demand) and reports that attention has a short-lived influence on performance of index of stocks, bonds and commodities. In addition, attention weakens the predictability of index return because more revealed information due to increasing attention improves market efficiency. We contribute to this stream of literature by showing investors' attention leads to larger shareholder base and improved stock liquidity.

The remainder of the paper is organized as follows. Section II describes the research design and the data. Sections III and IV present the empirical results. Section V describes the robustness checks. Section VI concludes by providing suggestions for future research.

2. Research Design and Data

2.1. Active Attention Measures

Since the beginning of 2004, Google Trends has provided data on the search frequencies of terms on a weekly basis (http://www.google.com/trends).[6] It shows how many searches have

[6] http://www.google.com/intl/en/trends/about.html. The data are scaled to the average traffic for the term in question over a fixed time period (usually January 2004).

been made for a specific keyword relative to the total number of searches over time.[7] Following Da et al. (2011) and Drake, Roulstone and Thornock (2012), we proxy investor attention by the search volume index (SVI) provided by Google Trends. Specifically, we measure investor attention for a company based on the SVI for the stock ticker rather than the company name, since searching for a stock using its ticker is less ambiguous (Da et al., 2011) and searches using ticker symbols as the search term are more likely to reflect searches for financial information than searches for non-financial information (Drake et al., 2012). We download the weekly SVI for the ticker symbols of S&P500 stocks, which provides time-series variations in the information searches for each firm. If a ticker is rarely searched for, Google Trends will return a zero value. In addition, we exclude two types of noisy tickers. First, we remove 12 companies whose tickers are single or double alphabets (e.g., "C" for Citi group, "M" for Macy's and "AA" for Alcoa). Second, we exclude 23 companies whose tickers have generic meanings (e.g., "DO" for Diamond Offshore Drilling, "GAS" for AGL resources, "LEG" for Legget & Platt and "FAST" for Fastenal).[8]

We download weekly SVIs for all constituents of the S&P 500 index over a six-year period from January 2004 to December 2009. A retail investor can easily obtain a firm's ticker from financial news, where tickers are often reported in parentheses. Following Da et al. (2011), we exclude SVIs with value of zero, and compute the change in SVI as follows:

[7] In this study, search is defined as the activity of submitting an enquiry regarding a particular term using Google. Consequently, the search volume is the number of enquiries submitted within a certain period.

[8] To confirm that the search of the tickers reflects retail investors' attention on the stocks, we employ a new application "Goolge Correlate" (http://www.google.com/trends/correlate), which identifies the most correlated SVIs. For example, the SVI of the ticker "APPL" is highly correlated with SVIs of "apple stocks" (correlation as 0.894), "apple quotes" (0.867) and "apple stock price"; while the SVI of "apple" is highly correlated with SVIs of with "apple store" (0.862), "iphone" (0.852) and "apple online store" (0.827). This indicates that investors tend to use tickers to search for stock-related information whereas consumers tend to search company's name for product and retail information, which justifies our strategy to use SVI for stock ticker instead of company name as a proxy for investors' attention.

$$\Delta SVI_t = Ln(SVI_t) - Ln[Med\ (SVI_{t-1}, \ldots\ldots, SVI_{t-8})] \qquad (1)$$

where SVI_t is the search volume index during week t obtained from Google Trends, and [$Med\ (SVI_{t-1}, \ldots\ldots, SVI_{t-8})$] is the median value of the SVI during the previous eight weeks. As a positive ΔSVI would indicate a surge in investor attention, a positive ΔSVI is more likely to lead to subsequent trading behavior. Another benefit of using ΔSVI is that time trends and low-frequency seasonality are removed (Da et al., 2011).

2.2. Passive Attention Measures

A commonly used passive attention measure is media coverage in newspapers. For example, Fang and Peress (2009) focus on four daily newspapers with nationwide circulation in the US: the *New York Times*, *USA Today*, the *Wall Street Journal*, and the *Washington Post*. We argue, however, that the average retail investor is unlikely to subscribe to more than two to three newspapers at the same time. A more convenient and inexpensive way for them to obtain news is through the internet, and every piece of news on the internet has "global circulation and access".

The advanced "news search" section in Google News enables us to obtain a figure for the total number of relevant news items per year, for each company in our sample, from 2004 to 2009.[9] To obtain the number of news items, we use the company name instead of the ticker, because tickers are only reported in financial newspapers but retail investors do not necessarily get their information from financial newspapers only. The multiple meanings of

[9] The number of news items related to a particular search term over a given period of time is available from Google News (http://news.google.com) database, which aggregates news from 4,500 English-language news sources worldwide. The stories are sorted without any consideration of political viewpoint or ideology.

the names of some companies may add noise to our data (e.g., Apple). However, due to the large number of news items, it would be unfeasible for us to read every article in order to exclude the irrelevant ones. Nevertheless, this noise is expected to introduce some bias against obtaining consistent results. A feature of the Google News is that it counts multiple newspaper distribution of the same article. Thus, it also reflects the dissemination of news, which is closely related to the passive attention of individuals.

Prior research also suggests that advertising expenditure is a measure of passive attention because intensive advertisement is able to promote the awareness of the product of the company among consumers and the stock of the company among investors (Grullon et al., 2004). In our study, we control for both Google News and firm's advertising expenditure so that the incremental effect of active attention reflected by SVI can be disentangled.

2.3. Research Design and Data

To investigate how investor attention affects the breadth of ownership and stock liquidity, we incorporate the attention measures to the models of Grullon et al. (2004) as follows:

$$\begin{aligned} LnNumS &= \lambda_0 + \lambda_1 SVI + \lambda_2 LnNews + \lambda_3 LnAdv \\ &+ \lambda_4 LnAge + \lambda_5 RET + \lambda_6 ROA + \lambda_7 LnMC + \lambda_8 (1/P_t) \\ &+ \lambda_9 LnTurnover + \lambda_{10} LnVolatility + \lambda_{11} NASDAQ + \varepsilon \end{aligned} \quad (2)$$

$$\begin{aligned} RBAS &= \gamma_0 + \gamma_1 SVI + \gamma_2 LnNews + \gamma_3 LnAdv \\ &+ \gamma_4 LnAge + \gamma_5 ROA + \gamma_6 LnMC + \lambda_7 (1/P_t) \\ &+ \gamma_8 LnTurnover + \gamma_9 LnVolatility + \gamma_{10} NASDAQ + \varepsilon \end{aligned} \quad (3)$$

where the number of shareholders (*lnNumS*) and the relative bid-ask spread (*RBAS*) are regressed against the search volume index (*SVI*), the number of news items available online

(*LnNews*), and the advertising expenses (*lnAdv*). To confirm our predictions, we expect to find a significantly positive λ_1 in equation (2) and a significantly negative γ_1 in equation (3). We also use the change in SVI (ΔSVI) instead of the level of the SVI for robustness checks. The change in SVI defined in Equation (1) reflects an abnormal "jump" in the SVI relative to the "normal" level over a longer time period (the previous eight weeks). As explained earlier, it can also remove time trends and low-frequency seasonality (Da et al., 2011). The annual observations of the number of shareholders, advertising expenses and other accounting data are obtained from Compustat. A large proportion of firms do not report their advertising expenses. Replacing missing advertising expenditure with zero is an approach commonly used in previous studies to maintain sample size (e.g., Grullon et al., 2004; Banker, Huang and Natarajan, 2011). In this study, because we use the natural logarithm of advertising expenditure in our analysis, we replace any missing values with \$0.01 rather than zero. As a robustness check, we also replicate the analysis based on the smaller sample excluding those firms with missing advertising expenditure and the results are consistent.

We calculate the relative bid-ask spread as the monthly average of the ratio of the daily inside spread to the midpoint of the daily inside spread from CRSP (Centre for Research in Security Prices). Chung and Zhang (2009) suggest the daily CRSP-based spread as a good substitute for the TAQ-based spread in that the former represents at least 91% (78%) of the cross-sectional variation in the latter from NASDAQ (NYSE/AMEX) stocks. We drop any observations of relative spread that are greater than 50% of the midpoint in order to filter the data for errors. We remove 26,732 daily observations with relative spreads larger than 50%, from the original 10,238,830 daily observations (accounting for 0.26%), by following Chung and Zhang (2009). We then transform the daily data into monthly data to perform the analysis. For robustness checks, we replicate the analysis using alternative liquidity measures,

including the effective spread and the relative effective spread. The change in relative spread is defined as the monthly change in relative spread in percentage terms. The effective spread is constructed as twice the difference between the transaction price and the spread midpoint. The relative effective spread is the effective spread scaled by the midpoint of the spread.

In order to perform the empirical analysis, we transform the daily liquidity spread measures and the weekly attention measures of SVI and ΔSVI to monthly observations by taking the average in each calendar month. Then, we merge the annual observations of the advertisement expense into the firm-month panel data.

Following Grullon et al. (2004), we control for other factors that may have an impact on stock liquidity. The market microstructure model (Ho and Stoll, 1980) suggests that a high trading volume reduces the inventory cost per trade and therefore leads to a smaller bid-ask spread. Hence, stocks with a high trading volume are expected to have smaller spreads. We control for share turnover (*LnTurnover*), which is constructed as the monthly average of the share volume divided by shares outstanding from CRSP. Large firms tend to have high trading volumes and thus smaller spreads, and therefore we also control for firm market capitalization (*lnMC*) from CRSP. Investors may have a preference for stocks within a certain price range, so we also include the inverse of the closing price from CRSP (*1/P*) in our analysis. Return volatility and firm age are included to proxy for risks. Return volatility is the monthly average of the standard deviation of daily returns, obtained from CRSP. Firm age is the number of years for which the firm has been included in CRSP. Average monthly return (*RET*) and return on assets (*ROA*) are used to control for market performance and profitability. Average monthly return is the average of the daily stock returns from CRSP. Return on assets is constructed from Compustat as the annual operating income before

depreciation, scaled by total assets. Finally, an exchange dummy (*NASDAQ* is assigned the value 1 for firms listed on the NASDAQ, and 0 otherwise) is included to account for systematic differences in the market microstructure. Following Grullon et al. (2004), for some variables we take their natural logarithm as shown in Equation (2) and (3), and we include industry and year fixed effects in the analysis[10]. The final sample consists of 14,690 firm-month observations over the period from 2004 to 2009. The top and bottom 0.5% of the variables are winsorized to reduce the possible effects of spurious outliers.

2.4. Summary Statistics

Panel A of Table 1 presents the descriptive statistics of the variables. Both the mean and median of the SVI change (ΔSVI) are positive, showing an upward trend in the attention paid to the tickers of S&P 500 firms. The media coverage of the firms, according to Google News, varies from 313 (25%) to 3,320 (75%), with a mean (median) of 9,914 (1,200). This is substantially larger than the amount of newspaper coverage documented in Fang and Peress (2009), where the mean (median) was 12 (5). The difference indicates that firms are better covered by online media than by traditional media such as national newspapers. The mean (median) advertising expenditure is $449 ($144) million, which is much larger than the figures documented in Grullon et al. (2004) based on an earlier sample from 1993 to 1998. This shows that firms are spending much more on advertising nowadays. The number of shareholders ranges from 4,000 to 51,000, with a mean (median) of 74,000 (14,000). The mean (median) of the relative spread is 0.029 (0.0241). The average firm in our sample is older and larger than that in Grullon et al. (2004), presumably for two reasons. First, we only include the constituents of the S&P 500, in which newly listed firms are less likely to be

[10] By following Grullon et al. (2004), the average monthly return (*RET*) is only incorporated in equation (2), that is, when the dependent variable is the number of shareholders (*lnNumS*).

included. Second, there is a threshold of search volume for Google Trends to report the SVIs, therefore SVIs are not available for some fledgling firms.

[Insert Table 1 about here]

2.5. The Active and Passive Attention Measures

In Table 2, we explore how the SVI and ΔSVI, the newly proposed direct measures of active attention, are related to the traditional passive attention measures, and to firm-specific characteristics. Table 2 shows that news coverage and advertising expenditure are positively associated with the SVI, which suggests that investors pay more attention to firms with greater visibility in terms of news coverage and expenditure on advertising. The coefficients of turnover, return on assets, firm size and return volatility are significantly positive, showing that firms with good operating performance, actively traded stocks, high market value, and higher risk grab more attention from retail investors. This is in line with the finding of Seasholes and Wu (2007) that stocks with higher returns or higher risk receive more news coverage and therefore attract more attention among investors. The coefficient of firm age is significantly negative, and this might be due to the impact of newly founded Information Technology glamour companies.[11] Despite the significance of the explanatory variables, the explanatory power of the model is low, and 95% of the variation in the active attention measures remains unexplained. In model II, we regress the change in SVI against the same set of explanatory variables. The only significant variable here is turnover, and more than

[11] We partition the sample according to the median of firm age, and run the regressions again on the two subsamples. We find that the coefficient of firm age is significantly positive ($p < 0.01$) in the subsample of older firms, and significantly negative ($p < 0.01$) in the subsample of younger firms. Since there are more Information Technology firms in the young subsample, we conjecture that the negative coefficient of firm age obtained for the full sample is attributed to their impact.

99% of the variation is unexplained. This shows the distinction in the aspects of attention captured by the active and passive measures.

[Insert Table 2 about here]

3. Active Attention Measure and Breadth of Ownership

We perform the regression analysis as shown in equation (2) to test the effect of investor attention on the shareholder base.[12] We regress the natural logarithm of the number of shareholders against the active attention measures, passive attention measures including online news coverage and advertising expenditure, and a set of control variables suggested in Grullon et al. (2004) to explain cross-sectional variations in the breadth of ownership. The results are reported in Panel A of Table 3. In model I we include only the SVI and the control variables. Here, the coefficient of SVI is significantly positive, which suggests that active attention is positively associated with the size of the shareholder base. The coefficients of firm age, firm size and return on assets are significantly positive, showing that profitable firms, large firms and long-standing firms enjoy a larger shareholder base. The coefficient of 1/P is positive and significant, in line with the explanation that individual investors are likely to buy stocks within a certain price range (i.e., higher 1/P).

We incorporate online news coverage in model II, and advertising expenditure in model III. The effect of the SVI remains significant after controlling for the passive attention measures. The coefficient of online news coverage is significantly positive, suggesting that firms that

[12] As shown in Panel A of the appendix, we first conduct a univariate test in the following way: we classify the firms into low-attention and high-attention subsamples based on the median value of the SVI. The former group of firms is associated with 47,310 fewer shareholders on average, and this difference is significant at the 1% level. We further classify the two subsamples into small and large firms, based on the median value of market capitalization. The difference between the number of shareholders for the low-attention and high-attention sub-samples is 2,270 and 74,380 for the small and large subsamples respectively, and the differences are statistically significant. The results support our prediction that firms with a higher amount of active attention paid to them will be associated with a larger shareholder base, no matter how large the firm is. The results of these tests are available from the authors upon request.

are widely covered by news stories on the internet are associated with a larger shareholder base. The impact of advertising expenditure is positive, but marginally insignificant. All active and passive measures are incorporated in model III, and the positive effect of the SVI on the shareholder base remains significant after controlling for the passive attention measures. The results are also economically significant. According to model I, a one standard deviation (1.34) increase in the SVI leads to an increase of 1,000 shareholders, which is 6% of the median number of shareholders (15,000) for our sample firms.

Because S&P 500 firms are observed multiple times in the firm-month panel data, we correct for standard errors using clustering in model IV by applying a bootstrapping regression as a robustness check. Standard errors are clustered by firm to account for heteroskedasticity. The coefficient of the SVI remains significantly positive. In model V, we replicate the test based on a smaller sample of 6,742 observations by excluding firms with missing advertising expenditure. The result shows that our main result hold for a subsample of firms with positive advertising expenditures[13]. When firm fixed effects are applied to control static firm-level effects in untabulated test, the results remain consistent. Overall, the results reported in Panel A show that the positive impact of active attention on the shareholder base is robust to the control of the passive attention measures, firm characteristics, and year and industry fixed effects.

In Panel B, we replicate the test by replacing the SVI with the change in SVI, and examine its impact on the shareholder base. Consistent with our prediction, the coefficient of the change in SVI is significantly positive in model I, showing that an increase in active attention leads to a larger number of shareholders. A one standard deviation (0.12) increase in the change of

[13] We use the full sample to perform the main test in that the subsample might be subject to selection bias. We thank the anonymous referee for this point.

SVI leads to an increase of about 1,000 in the shareholder base, which is about 6% of the corresponding figure (15,000) for a median firm. The positive impact, again, is robust to the control of the passive attention measures, firm characteristics, and year and industry fixed effects, as shown in models II and III. The signs of the coefficients for online news and advertising expenditure and the other control variables are consistent with those reported in Panel A. Finally, we adjust for standard errors and apply the bootstrapping regression in model IV as a robustness check, and replicate the test based on a smaller sample of 6,742 firm-month observations excluding observations with missing advertising expenditure in model V. The findings remain consistent. To sum up, the results reported in Table 3 show that retail investors tend to become shareholders of the listed firms to which they pay attention through internet searches. The results suggest that the internet has become an important tool for retail investors to gather information and make investment decisions.

[Insert Table 3 about here]

4. Active Attention and Stock Liquidity

Table 4 reports the results of the impact of investor attention on the stock liquidity. As shown in Equation (3), we regress the relative bid-ask spread on the SVI, online media coverage, advertising expenditure and a set of control variables.[14] In model I, we include only the SVI and the control variables. Capturing the impact of active attention, the coefficient of the SVI is significantly negative, which suggests that higher level of investor attention reflected by

[14] As shown in the Panel B of the Appendix, we divide our sample firms into low-attention and high-attention subsamples based on the median level of the SVI and test the difference in the means of the relative bid-ask spread. The results show that the relative bid-ask spread is significantly smaller in the high-attention subsample. When we further divide the sub-samples according to market capitalization, into small and large firms, the difference in the bid-ask spread still exists in both types of firm. The difference is more pronounced for smaller firms because they are, in general, less recognized by investors, and therefore more likely to benefit from increased active attention from investors. The results are available upon request.

search frequency leads to a reduced bid-ask spread and therefore improved stock liquidity. Model II and Model III incorporate passive attention measures including Google online news coverage and advertising expenditure, while Model IV adjusts the clustered standard error by firm and apply the bootstrapping regression. Model V is based on the reduced sample of 6,742 observations with non-missing advertising expenditure. The coefficients of the SVI remain negative and significant across the models, suggesting that investors' active attention helps to improve stock liquidity. Consistent with Grullon et al. (2004), the coefficient of advertising expenditure is significantly negative in Model V.

We also replicate the analysis by replacing the SVI with the change in SVI, and present the results in Panel B of Table 4. In model I, the coefficient of the SVI change is significantly negative, suggesting that an increase in investor attention improves liquidity. The results remain robust after controlling for the passive attention measures in Models II and III, after applying the bootstrapping regression model to adjust the clustered standard error in Model IV and after dropping the observations with no advertisement expenditure in Model V. In untabulated test, we also apply firm fixed effects or replace the level of liquidity with the change in liquidity[15] as the dependent variable. The results remain consistent in that the significantly negative coefficients of the change in SVI support our prediction that increased investor attention helps to promote stock liquidity[16].

There are several aspects of liquidity, and bid-ask spread reflects the inventory aspect. Stoll (1978) and Ho and Stoll (1981) argue that liquidity depends on factors that influence the risk

[15] The change in liquidity is measured by the relative spread in month *t* minus the average relative spread between month *t-1* and month *t-3*.

[16] The panel C in appendix also reports the results by applying the firm fixed effects. The results of three out of four regression models are broadly in line with our main results. Petersen (2006) gives possible explanations for the minor inconsistency: the standard errors clustered by firm are unbiased and produce correctly sized confidence intervals whether the firm effect is permanent or temporary while the fixed effect and random effects model only produces unbiased standard errors when the firm effect is permanent.

of holding inventory and extreme events that provoke order imbalance and thereby cause inventory overload. To compensate the inventory holding cost, dealers will purchase shares at the bid price below the "true" price and sell shares at the ask price above the "true' price, generating the bid-ask spread.

Bid-ask spread, however, may fail to capture other aspects. For example, Grossman and Miller (1988) show that liquidity is also determined by the demand and supply of immediacy, and bid-ask spread does not reflect the cost of supplying immediacy to the market. Kyle (1985) notes the informed speculation aspect of liquidity that market makers cannot distinguish between order flow generated by informed traders or liquidity traders, they set the price as an increasing function of the imbalance in the order flow, which may indicate informed trading. This suggests a positive relation between the transaction volume and price change, known as price impact. It can be considered as the first derivative of the effective spread with respect to the order size and reflects the cost of demanding additional instantaneous liquidity. Amihud (2002) measure it as the daily price response associated with one dollar of trading volume, calculated as the daily ratio of absolute stock return to dollar trading volume averaged over all positive volume days.

To explore whether the active attention of retail investors also affect other aspects of liquidity, we replace the dependent variable as the Amihud (2002) liquidity measure (*ILLIQ*) and the results are reported in Panel C. The coefficient of Change in SVI is insignificant. When we incorporate *ILLIQMA*, estimated as the ratio of the *ILLIQ* to the average *ILLIQ* of all stocks in the market, or incorporate the standard deviation of *ILLIQ* (Lang and Maffett, 2011), the coefficients of attention measures remain insignificant. The results suggest that the active attention could significantly mitigate the adverse-selection type of illiquidity, but not

the price-impact type of illiquidity. Amihud (2002) also note that although bid-ask spread is a finer and better measure, there is no one single measure that captures all its aspects. As suggested by Da et al (2011), SVI largely reflects the attention of retail investors because institutional investors have access to professional information vendors. In general the trading volume (in dollar term) of an average retail investor is less likely to be large, which suggests that their trading behavior might have diminished effect on the stock price reaction to trading.

[Insert Table 4 about here]

5. Robustness Checks

5.1. Alternative Bid-Ask Spread Measures

To confirm that our finding is robust to other liquidity measures, we replicate the regressions by replacing the relative bid-ask spread with the effective spread and the relative effective spread as the dependent variables. Following Grullon et al. (2004), the effective spread is defined as twice the difference between the transaction price and the spread midpoint, and the relative effective spread is defined as the effective spread divided by the midpoint of the spread. Table 5 provides the results of the analysis. The change in the SVI is significantly negative in both model I and model II of Panel A. This is consistent with our main finding that active attention helps to improve stock liquidity. Both online news coverage and advertising expenditure are found to reduce the bid-ask spread as well. Next, we use the turnover rate as an alternative liquidity measure and repeat the analysis (see Datar et al., 1998).[17] In Panel B, we show that both the active and passive attention measures are

[17] Trading activity reflected by turnover rate is a weak measure of liquidity because trading volume could be high when liquidity is low (Pastor and Stambaugh, 2003). Grullon et al. (2004) find the turnover rate to be weakly correlated with advertising expenditure.

positively associated with the share turnover. This shows that our results are robust to various specifications of liquidity.

[Insert Table 5 about here]

5.2. Propensity Score Matching

Our research design may be subject to endogeneity concern. The concern stems from possible reverses causality or selection effects in that retail investor attention is not randomly assigned to firms. For example, liquidity shocks may affect retail investor attention. In addition, corporate announcements, events or firm characteristics may affect both liquidity level and retail investor attention. To control for the endogeneity issue and to draw causal inferences, and to explore attributors of substantial increase in retail investor attention, we apply propensity score matching (PSM) (Rosenbaum and Rubin, 1983) to replicate our main tests. To implement the approach, we firstly define substantial increase in SVI (*SIn_SVI*) is equal to 1 if the change in SVI is above the 10% percentile and 0 otherwise. We then estimate a logit regression to the incidence of substantial increase in SVI (*SIn_SVI*) based on traditional attention measures and firm characteristics as shown in Table 5D. Then, we construct a one-to-one nearest-neighbor matched sample based on the closest predicted value (propensity scores) from the logit regression. In this way, each firm with substantial increase in attention (*SIn_SVI*=1) is matched with another counterpart firm with similar characteristics, but without substantial increase in attention (*SIn_SVI*=0). It is worth to note that the results of Model I that firms with more advertising costs, larger size, higher operating performance, higher turnover, and higher volatility, and firms listed in NASDAQ are more likely to experience substantial increase in SVI, whereas firms which also attract high media coverage is less likely to experience such change. Such randomized experiment sample with 3002 observations is less subject to the endogeneity issue. As shown in Table 5C, we

replicate the main tests by using the matching sample to the more reliable and rigorous results. The results remain consistent that substantial increase in SVI improves breadth of ownership and stock liquidity.

To control possible confounding effects of corporate events, we replicate the PSM test by incorporating dummy variables capturing index inclusion, corporate splits or dividend payments. It is also worth to note that these corporate events do not significantly affect the change in attention as documented in the first step of the PSM approach. The results from the second step remain consistent.

5.3. Test for Causality

To further strengthen the causal inferences of the results, we also employ the Granger test as additional robustness checks by following the studies of the determinants of stock liquidity (e.g. Roulstone, 2003; Chordia et al., 2005; and Goyenko et al., 2009). [18] The Granger causality test is used to determine whether one time series is useful in forecasting another. The logic is as follows: suppose that we have three time series: the change in investor attention (ΔSVI), the liquidity (RBAS) and a vector of control variables that have predictive power for liquidity (Control). We first use past values of RBAS (up to previous 10 weeks), ΔSVI and Control to forecast RBAS (equation 4 below). Then, we use past values of ΔSVI, RBAS and Control to predict ΔSVI. If the results reject the hypothesis that past values of RBAS can predict ΔSVI (equation 5 below) but fail to reject the hypothesis that past values of ΔSVI can predict RBAS (equation 4), this will indicate that the past values of ΔSVI

[18] For example, Roulstone (2003) uses Granger causality test to identify the causal relationship between analyst following and market liquidity. His results show that analyst following causes stock liquidity to improve, because post levels of analyst following are related to future liquidity levels.

provide statistically significant information about RBAS. That is, ΔSVI is able to Granger-cause RBAS.

First, in order to check whether changes in the SVI lead to changes in liquidity, we use a standard F test to test the joint hypothesis that $\delta_1 = \delta_2 = ...\delta_{10} = 0$ for the following regression:

$$RBAS_{i,t} = \alpha + \sum_{j=1}^{10} \beta_j RBAS_{i,t-j} + \sum_{k=1}^{10} \delta_k \Delta SVI_{i,t-k} + \sum Control \qquad (4)$$

Then, we test the joint hypothesis that $\kappa_1 = \kappa_2 = ...\kappa_{10} = 0$ for the following regression:

$$\Delta SVI_{i,t} = \sigma + \sum_{j=1}^{10} \kappa_j RBAS_{i,t-j} + \sum_{k=1}^{10} \gamma_k \Delta SVI_{i,t-k} + \sum Control \qquad (5)$$

The untabulated results show the hypothesis that a change in the SVI does not cause a change in liquidity is rejected at the 10% confidence level. However, we fail to reject the hypothesis that liquidity does not cause a change in the SVI at a conventional level. The findings stay robust to the inclusion of higher-order lags of liquidity. Therefore, our inference is that an increase in investors' attention paid to a stock (reflected by a change in the SVI) causes liquidity to improve.

5.4. Other confounding factors

To show that our results are driven by confounding corporate events such as stock split and dividends, we incorporate them as additional control variables in our tests. The first dummy variable is *Event_split* that equals 1 if share splits happened in the month and 0 otherwise; similarly, *Event_dividend* equals 1 if the dividend distribution was announced in the month and 0 otherwise. Alternatively, we also replicate the tests by excluding the firm-month

observations with these corporate events, and the results are robust. The results reported in the columns I to IV in Table 4D remain consistent in that investor attention measured by Change in SVI increases shareholder base and improve stock liquidity. The results also hold when we incorporate other events such as *Index*, which equals 1 if the firm is newly included in S&P500 index in the month and 0 otherwise.

As previous literature shows that financial reporting transparency is associated with stock liquidity (i.e. Lang and Maffett, 2011), we create two dummy variables to proxy for financial reporting transparency. Specifically, *Big4* is set to 1 for firms audited by Big four auditing firms (namely KPMG, Ernst & Young, Deloitte Touche Tohmatsu, and PricewaterhouseCoopers), and 0 otherwise. We add them as additional control variables in the analysis as reported in column V of Table 5D, and find that the results remain consistent. We also consider alternative measures for financial reporting transparency such as going on concern audit opinions from auditing firm but find that S&P500 firms in our sample do not receive going on concern audit opinions.

6. Conclusion

The "home bias" literature suggests that investors are inclined to invest in firms that they are familiar with. In order to get familiar with a firm, investors need to acquire relevant information. Individual investors, who are generally unable to access professional information vendors such as Reuters or Bloomberg, may increasingly rely on Google, the dominant Internet search engine, to search for relevant information before making investment decisions. In this paper, we use search frequency data on S&P 500 stocks between January 2004 and December 2009, provided by Google Trends, as a direct measure of active investor attention, and examine the impact of this active attention from retail investors on the

shareholder base and stock liquidity. We find this active attention measure to be distinct from passive measures such as the number of news items available on the Internet (based on Google News; news.google.com) [19] and advertising expenditure. Despite the positive correlation between the active and passive attention measures, almost 95% of the cross-sectional variation in the former cannot be explained by the latter. We further show that the increased investor attention indicated by the search volume index (SVI) and Google News contributes to a broader shareholder base. Furthermore, increased investor attention leads to a reduced relative bid-ask spread and a higher turnover rate. Our findings are robust to the control of firm characteristics suggested in Grullon et al. (2004), and to alternative measures of stock liquidity.

Our study contributes to the burgeoning literature on the role of investor attention in the dynamics of asset pricing. Studies in this stream of literature include Barber and Odean (2008) on investor attention and individual investors' trading behavior, Yuan (2009) on recording-breaking events related to the Dow index and front-page coverage in newspapers as proxies for investor attention, and its impact on trading behavior and market returns, and Da et al. (2011) on investor attention measured by Google search frequency, and its effect on IPO returns and the price pressure hypothesis proposed by Barber and Odean (2008). This study also extends the literature on the "investor recognition hypothesis" (e.g., Grullon et al., 2004; Fang and Peress, 2009). In markets with information asymmetry, investors are less likely to possess the required information. Consequently, securities with lower investor recognition become less liquid and have to offer a higher return to compensate for their

[19] Yuan (2009), Tetlock (2010) and Fang and Peress (2009) use either the LexisNexis database or the Dow Jones news archive to determine the number of newspaper articles related to a stock. The Google news channel includes news from the most popular English-language news sites, such The New York Times, Bloomberg, Reuters, the Guardian, CBS News, BBC News, Times Online, and CNN, and thus offers broader news coverage. We argue that online media coverage is more accessible than newspaper coverage for retail investors, who are more likely to search for information on the internet.

"illiquidity". The fact that a security is attracting more attention from investors can enable it to be "better recognized". As a result, stocks with increased investor attention become more liquid. Our results generally lend credence to the "investor recognition" hypothesis.

The findings of our study have implications for companies that wish to promote investor recognition. Companies may intentionally make themselves more visible on the Internet and especially in the Google search engine in order to attract the attention of potential investors. Our results might be of interest to participants in the financial markets (e.g., liquidity traders), in that they may benefit from sophisticated models that incorporate individuals' information acquisition behavior into predictions of stock liquidity. Finally, our findings may incentivize Google and other search engine companies to further improve their service in terms of providing more timely and accurate data on public search behavior, and so increase their profits from such services.

Our study is subject to the following limitations. First, although we apply propensity score matching and Granger test to address the concern of endogeneity problem, future research may attempt to identity some exogenous shift in investor's attention and further explore the casual relation between investors' attention and stock liquidity. We notice that corporate events could not trigger substantial change in retail investors' attention in that non-professional investors do not tend to closely follow corporate announcements. We conjecture that retail investor attention could be affected by proxies related to social media. Second, our analysis is exclusively built on the sample of S&P 500 stocks, which are large and well-established firms with high visibility. It is expected that smaller and less recognized firms would benefit more from increased investor attention, therefore adding small firms to future studies might strengthen their analysis. Note that if the tickers of some small firms are rarely

searched for, Google will return a value of zero. Third, as suggested by Da et al (2011), SVI largely reflects the attention of retail investors because institutional investors have access to professional information vendors such as the Bloomberg or Reuters. In general the trading behavior of an average retail investor is less likely to have a remarkable effect on the stock price, which might plausibly explain why active attention measure of retail investors is insignificantly related to the Amihud (2002) illiquidity measure. Further research may examine a wide range of liquidity measures. Finally, the majority of research in this area is based on US data.[20] As Google becomes an increasingly important source of information for investors around the world, it might be interesting to explore the capital market consequences of investors' active demand for information in other markets (e.g., the UK, other European countries and Asia). We leave this for future research.

[20] One exception is Aouadi et al. (2013), which construct attention measure with French investor's online search behavior, provided by Google, and show that investor attention is a determinant of the stock market liquidity and volatility.

Reference

Amihud, Y. 2002. Illiquidity and stock returns: cross-section and time-series effects, Journal of Financial Market, 5, 31-56.

Amihud, Y., and Mendelson, H. 1980. Dealership market: market making with inventory. Journal of Financial Economics, 8, 31-53.

Aouadi, A., Arouri, M., and Teulon, F. 2013. Investor attention and stock market activity: Evidence from France. Economic Modelling, 35, 674-681.

Banker, R.D., Huang, R., and Natarajan, R. 2011. Equity incentives and long-term value created by SG&A Expenditure. Contemporary Accounting Research 28, 794-830.

Barber, B., and Odean, T. 2008. All that glitters: The effect of attention and news on the buying behaviour of individual and institutional investors. Review of Financial Studies 21, 785-818.

Bushee, B, and Miller, G. 2012. Investor relations, firm visibility and investor following. The Accounting Review 87, 867-897.

Cao, H., Han, B., Hirshleifer, D., and Zhang, H. 2011, Fear of the unknown: Familiarity and economic decisions. Review of Finance 15, 173-206.

Chen, H., Noronha, G., and Singal, V. 2004, The price response to S&P 500 index additions and deletions: evidence of asymmetry and a new explanation. Journal of Finance 59, 1901-1930.

Chordia, T., Sarkar, A., and Subrahmanyam, A. 2005. An empirical analysis of stock and bond market liquidity. Review of Financial Studies 18, 85-129.

Chung, K.H., and Zhang, H. 2009. A simple approximation of intraday spreads using daily data. Working paper, State University of New York at Buffalo. Available at http://ssrn.com/abstract=1346363.

Da, Z., Engelberg, J., and Gao, P. 2011. In search of attention. Journal of Finance 66, 1461-1499.

Datar, V., Naik, N., and Radcliffe, R. 1998. Liquidity and stock return: an alternative test. Journal of Financial Market 1, 203-29.

Drake, M., Roulstone, D., and Thornock, J. 2012. Investor information demand: Evidence from Google searches around earnings announcement. Journal of Accounting Research 50, 1001-1040.

Engelberg, J., and Parsons, C. 2011. The casual impact of media in financial markets. Journal of Finance 66, 67-97.

Fang, L.H., and Peress, J. 2009. Media coverage and the cross-section of stock returns. Journal of Finance 64, 2023-2052.

French., and Poterba, J. 1991. Investor diversification and international equity markets. American Economic Review 81, 222-226.

Gao, L., Li, O.Z., and Eric Yeung, E. 2012. Online information acquisition and investor trading. Working paper, University of Georgia. Available at http://ssrn.com/abstract=1929402.

Grossman, S., and Miller, M. 1988. Liquidity and market structure. Journal of Finance 43, 617-633.

Grullon, G., Kanatas, G., and Weston, J.P. 2004. Advertising, breadth of ownership, and liquidity. Review of Financial Studies 17, 439-461.

Goyenko, R.Y., and Ukhov, A.D. 2009. Stock and bond market liquidity: A long-run empirical analysis. Journal of Financial and Quantitative Analysis 44, 189-212.

Hirshleifer, D., and Teoh, S.H. 2003. Limited attention, information disclosure, and financial reporting. Journal of Accounting and Economics 36, 337-386.

Hirshleifer, D., Lim, S., and Teoh, S.H. 2011. Limited investor attention and stock market mis-reactions to accounting information. Review of Asset Pricing Studies 1, 35-73.

Ho, T., and Stoll, H.R. 1980. On dealer markets under competition. Journal of Finance 35, 259-267.

Ho, T., and Stoll, H.R. 1981. Optimal dealer pricing under transaction and return uncertainty. Journal of Financial Economics 9, 47-73.

Hou, K., Peng, L., and Xiong, W. 2008. A tale of two anomalies: The implications of investor attention for price and earnings momentum. Working paper, Ohio State University and Princeton University.

Kyle, A. 1985. Continuous auction and inside trading. Econometrica 53, 1315-1336.

Lang, M., and Maffett, M. 2011. Transparency and liquidity uncertainty in crisis period. Journal of Accounting and Economics 52, 101-125.

Lehavy, R., and Sloan, R. 2008. Investor recognition and stock returns. Review of Accounting Studies 13, 327-362.

Merton, R. 1987. A simple model of capital market equilibrium with incomplete information. Journal of Finance 42, 483-510.

Pastor, L., and Stambaugh, R.F. 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642-685.

Petersen, M. 2006. Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. Kellogg Finance Dept. Working Paper No. 329; AFA 2006 Boston Meetings Paper. Available at SSRN: http://ssrn.com/abstract=661481

Rosenbaum, P., and Rubin, D. 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41-55.

Roulstone, D. 2003. Analyst following and market liquidity. Contemporary Accounting Research 20, 551-578.

Seasholes, M., and Wu, G. 2007. Predictable behavior, profits, and attention. Journal of Empirical Finance 14, 590-610.

Siganos, A. 2013. Google attention and target price run ups. International Review of Financial Analysis 29, 219-226.

Stoll, H.R. 1978. The pricing of security dealer service: an empirical study of Nasdaq stocks. Journal of Finance 33, 1153-1172.

Tesar, L., and Werner, I. 1995. Home bias and high turnover. Journal of International Money and Finance 14, 467-492.

Tetlock, P.C. 2010. Does public financial news resolve asymmetric information? Review of Financial Studies 23, 3520-2557.

Vlastakis, N., and Markellos, R. 2012. Information demand and stock market volatility. Journal of Banking and Finance 36, 1808-1821.

Vozlyublennaia, N. 2014. Investor attention, index performance and return predictability. Journal of Banking and Finance, Forthcoming.

Yuan, Y. 2009, Attention and trading. Working paper, University of Iowa.

Table 1. Summary Statistics

	Mean	Std. Dev	25%	Median	75%	Obs.
Investor Attention						
SVI	1.1752	1.3412	0.8000	0.9800	1.2200	14,690
SVI change (ΔSVI)	0.0079	0.1200	-0.0400	0.0013	0.0470	14,690
number of news	9,914	49,515	313	1200	3,320	14,690
Advertising(million)	511	729	50	212	629	6,742
Advertising'(million)	235	556	.01	.01	163	14,690
Breadth of Ownership						
number of Shareholders (thousand)	74	199	4	15	51	14,690
Liquidity Measures						
Relative Spread	0.0297	0.0186	0.0181	0.0241	0.0347	14,690
Firm Characteristics						
firm age	35	24	15	34	47	14,690
Stock return	0.0098	0.1000	-0.0400	0.0100	0.0600	14,690
ROA	0.5700	0.5000	0.3000	0.4400	0.6600	14,690
Firm size	9.3300	1.1500	8.5100	9.1700	10.0800	14,690
1/Share price	0.0400	0.0400	0.0200	0.0300	0.0400	14,690
Log (return volatility)	-4.0200	0.5600	-4.4200	-4.0800	-3.6800	14,690
NASDAQ	0.1500	0.3600	0	0	0	14,690

This table reports the summary statistics of our sample. The sample includes the constituents of the S&P 500 over a period of six years from January 2004 to December 2009. The SVI is the search frequency of a stock ticker, from Google Trends. The change in the SVI is the difference between the SVI during week *t* and the median value of the SVI during the previous eight weeks. Number of news is the online media coverage from Google News. Advertising expenses and number of shareholders are obtained from Compustat. Note that not all firms disclose their advertising expenditure, and the number of firm-year observations is only 6,742. In order to keep as many observations as possible in our analysis, we replace the missing observations with 0.01, assuming these firms spend roughly zero on advertising, following Grullon et al. (2004), and this is expressed as *Advertising'*. The relative bid-ask spread is the monthly average of the ratio of the daily inside spread to the midpoint of the daily inside spread from CRSP. Firm age is the number of years for which the firm has been included in CRSP. The average monthly return is the average of the daily stock returns from CRSP. Return on assets is constructed from Compustat as the annual operating income before depreciation, scaled by total assets. Firm size is the market capitalization, calculated as the product of the total number of outstanding shares and the annual closing price. Share turnover is constructed from CRSP as the monthly average share volume divided by the shares outstanding. Return volatility is the monthly average of the standard deviation of daily returns, obtained from CRSP. NASDAQ equals 1 for firms listed on the NASDAQ, and 0 otherwise.

Table 2. Active and Passive Attention Measures

	Dependent Variables	
	Model I: SVI	Model II : Change in SVI
Ln (number of news)	0.013***	-0.0004
	(6.29)	(-0.42)
Ln (advertising)	0.008***	0.0002
	(7.64)	(1.16)
Ln (turnover)	0.08***	0.0004**
	(9.13)	(2.2)
Ln (firm age)	-0.03***	-0.0004
	(-4.24)	(-0.35)
ROA	0.07***	-0.0001
	(6.71)	(-0.06)
Ln (Firm size)	0.08***	0.005***
	(14.5)	(5.02)
Ln (return volatility)	0.03***	0.007***
	(2.65)	(2.75)
Obs.	14,690	14,690
Adj. R^2	0.05	0.003

This table shows to what extent the active attention (measured by SVI and the change in SVI) can be explained by the passive attention (measured by online media coverage and advertising expenditure), and firm characteristics. The SVI is the search frequency of the stock ticker, according to Google Trends. The change in the SVI is the difference between the SVI during week t and the median value of the SVI during the previous eight weeks. The "number of news" is the online media coverage, according to Google News. Advertising expenditure is obtained from Compustat. Firm age is the number of years for which the firm has been included in CRSP. Return on assets is constructed from Compustat as the annual operating income before depreciation, scaled by total assets. Firm size is the market capitalization, calculated as the product of the total number of outstanding shares and the annual closing price. Return volatility is the monthly average of the standard deviation of daily returns, taken from CRSP. *, ** and *** represent significance at the 10%, 5% and 1% levels. Standard errors are adjusted for heteroscedasticity.

Table 3. Active Attention and Breadth of Ownership: Multivariate Analysis

This table reports the estimates from the panel regressions relating the number of shareholders to the active investor attention (SVI). The sample includes the constituents of the S&P 500 over a period of six years from January 2004 to December 2009. The SVI is the search frequency of a stock ticker, obtained from Google Trends. The change in the SVI is the difference between the SVI in week *t* and the median value of the SVI during the previous eight weeks. The "number of news" is the online media coverage obtained from Google News. Advertising expenses and the number of shareholders are taken from Compustat. Firm age is the number of years for which the firm has been included in CRSP. Average monthly return is the average of the daily stock returns from CRSP. Return on assets is constructed from Compustat as the annual operating income before depreciation, scaled by total assets. Firm size is the market capitalization, calculated as the product of the total number of outstanding shares and the annual closing price. Share turnover is constructed from CRSP as the monthly average of the share volume divided by the shares outstanding. Return volatility is the monthly average of the standard deviation of daily returns, drawn from CRSP. NASDAQ equals 1 for firms listed on the NASDAQ, and 0 otherwise. *, ** and *** represent significance at the 10%, 5% and 1% levels. Standard errors are adjusted for heteroscedasticity.

Panel A. SVI and Breadth of Ownership

	Dependent Variable				
	Ln (Number of Shareholders in thousand)				
	Model I	Model II	Model III	Model IV	Model V
SVI	0.02**	0.02**	0.02**	0.19**	0.21**
	(2.36)	(2.39)	(2.33)	(7.92)	(4.66)
Ln (number of News)		0.04***	0.04**	0.02**	0.04***
		(5.51)	(5.42)	(2.41)	(5.51)
Ln (advertising)			0.01	-0.01	0.26***
			(1.57)	(-0.74)	(14.69)
Ln (firm age)	0.27***	0.27***	0.27***	0.12***	0.34***
	(12.2)	(12.25)	(12.25)	(8.07)	(7.1)
Return	-0.2	-0.2	-0.2	-0.14	0.14
	(-1.59)	(-1.58)	(-1.56)	(-1.09)	(0.86)
ROA	0.31***	0.31***	0.31***	0.03	0.83***
	(8.66)	(8.77)	(8.8)	(0.54)	(7.56)
Ln (Firm size)	0.90***	0.88***	0.87***	0.97***	0.78***
	(73.06)	(66.07)	(64.4)	(41.41)	(17.64)
1/Share price	8.28***	8.18***	8.16***	11.53***	14.15***
	(13.09)	(12.97)	(12.96)	(13.22)	(12.08)
Ln (turnover)	-0.07**	-0.08**	-0.08**	-0.26***	-0.22**
	(-2.47)	(-2.56)	(-2.62)	(-5.89)	(-2.69)
Ln (return volatility)	-0.13***	-0.14***	-0.13***	-0.10***	0.03
	(-3.74)	(-3.80)	(-3.72)	(-2.63)	(0.61)
NASDAQ	-0.67***	-0.65***	-0.65***	-0.77***	-0.46***
	(-13.93)	(-13.55)	(-13.53)	(-10.30)	(-5.79)
Year	Y	Y	Y	Y	Y
Industry	Y	Y	Y	Y	Y
Adjust clustered errors	N	N	N	Y	Y
Exclude missing advertising	N	N	N	N	Y
Obs.	14690	14690	14690	14690	6742
Adj. R^2	0.397	0.398	0.398	0.28	0.34

Panel B. Change in SVI and Breadth of Ownership

	Dependent variable				
	Ln (Number of Shareholders in thousand)				
	Model I	Model II	Model III	Model IV	Model V
Change in SVI	0.21**	0.21**	0.21**	0.37**	0.22**
	(1.93)	(1.96)	(1.96)	(2.53)	(1.88)
Ln (number of News)		0.04***	0.04**	0.03***	-0.01
		(5.52)	(5.42)	(2.94)	(-0.69)
Ln (advertising)			0.01	-0.01	0.30***
			(1.64)	(-0.32)	(17.83)
Ln (firm age)	0.27***	0.27***	0.27***	0.13***	0.31***
	(12.15)	(12.2)	(12.2)	(6.49)	(6.63)
Return	-0.21	-0.21	-0.2	-0.14	0.1
	(-1.64)	(-1.63)	(-1.60)	(-1.53)	(0.79)
ROA	0.31***	0.31***	0.31***	-0.02	0.80***
	(8.59)	(8.7)	(8.73)	(-0.35)	(8.19)
Ln (Firm size)	0.90***	0.88***	0.88***	0.97***	0.80***
	(72.68)	(66.7)	(64.1)	(51.57)	(25.95)
1/Share price	8.24***	8.14***	8.12***	11.02***	14.83***
	(13.13)	(13)	(13)	(20.76)	(20.61)
Ln (turnover)	-0.07**	-0.08**	-0.08**	-0.26***	-0.13**
	(-2.47)	(-2.57)	(-2.64)	(-6.92)	(-1.82)
Ln (return volatility)	-0.14***	-0.14***	-0.14***	-0.10**	0.02
	(-3.83)	(-3.90)	(-3.81)	(-2.53)	(0.4)
NASDAQ	-0.67***	-0.65***	-0.65***	-0.76***	-0.52***
	(-13.95)	(-13.56)	(-13.55)	(-11.23)	(-8.42)
Year	Y	Y	Y	Y	Y
Industry	Y	Y	Y	Y	Y
Adjust cluster errors	N	N	N	Y	Y
Exclude missing advertising	N	N	N	N	Y
Observation	14,690	14,690	14,690	14,690	6,742
Adj. R^2	0.397	0.398	0.398	0.28	0.34

Table 4. Active Attention and Stock Liquidity

This table reports the estimates from panel regressions relating the relative bid-ask spread to active investor attention (SVI). The sample includes the constituents of the S&P 500 over a period of six years from January 2004 to December 2009. The SVI is the search frequency of the stock ticker, obtained from Google Trends. The change in the SVI is the difference between the SVI in week *t* and the median value of the SVI during the previous eight weeks. The "number of news" is the online media coverage from Google News. Advertising expenses are obtained from Compustat. The relative bid-ask spread is the monthly average of the ratio of the daily inside spread to the midpoint of the daily inside spread, obtained from CRSP. Firm age is the number of years for which the firm has been included in CRSP. Return on assets is constructed from Compustat as the annual operating income before depreciation, scaled by total assets. Firm size is the market capitalization, calculated as the product of the total number of outstanding shares and the annual closing price. Share turnover is constructed from CRSP as the monthly average of the share volume, divided by shares outstanding. Return volatility is the monthly average of the standard deviation of daily returns, from CRSP. NASDAQ equals 1 for firms listed on the NASDAQ, and 0 otherwise. *, ** and *** represent significance at the 10%, 5% and 1% levels. Standard errors are adjusted for heteroscedasticity.

Panel A. SVI and relative bid-ask spread

	Dependent Variable Relative spread in %				
	Model I	Model II	Model III	Model IV	Model V
SVI	-0.01**	-0.01**	-0.01***	-0.01*	-0.01**
	(-2.18)	(-2.17)	(-2.11)	(-1.73)	(-1.87)
Ln (number of News)		0.01*	0.01*	0.01	0.01
		(1.71)	(1.8)	(1)	(0.15)
Ln (advertising)			-0.003	-0.01	-0.02***
			(-1.42)	(-0.34)	(-2.56)
Ln (firm age)	-0.01	-0.01	-0.01	-0.02***	-0.04***
	(-0.73)	(-0.69)	(-0.70)	(-2.44)	(-3.72)
ROA	-0.01	-0.01	-0.01	-0.03***	-0.12***
	(-0.46)	(-0.41)	(-0.44)	(-2.72)	(-5.88)
Ln (firm size)	0.003	-0.001	0.001	0.01	0.03***
	(0.39)	(-0.17)	(0.16)	(0.43)	(4.22)
1/Share price	2.95***	2.93***	2.94***	3.17***	3.64***
	(8.04)	(7.99)	(7.99)	(11.58)	(9.77)
Ln (turnover)	0.12***	0.12***	0.12***	0.15***	0.16***
	(7.68)	(7.64)	(7.69)	(13.24)	(8.32)
Ln (return volatility)	2.47***	2.47***	2.47***	1.93***	1.82***
	(84.99)	(84.95)	(84.86)	(87.23)	(38.39)
NASDAQ	-0.03	-0.03	-0.03	0.02	0.01
	(-1.28)	(-1.14)	(-1.14)	(1.43)	(0.11)
Year	Y	Y	Y	Y	Y
Industry	Y	Y	Y	Y	Y
Adjust cluster errors	N	N	N	Y	Y
Exclude missing advertising	N	N	N	N	Y
Observation	14,690	14,690	14,690	14,690	6,742
Adj. R^2	0.793	0.793	0.793	0.56	0.55

Panel B. Change in SVI and relative bid-ask spread

	Dependent Variable Relative spread in %				
	Model I	Model II	Model III	Model IV	Model V
Change in SVI	-0.24***	-0.24***	-0.24***	-0.19***	-0.27***
	(-4.30)	(-4.30)	(-4.30)	(-4.58)	(-6.45)
Ln (number of News)		0.01*	0.01*	0.01	0.01
		(1.7)	(1.79)	(0.88)	(0.12)
Ln (advertising)			-0.003	-0.01	-0.01**
			(-1.42)	(-0.16)	(-2.47)
Ln (firm age)	-0.01	-0.01	-0.01	-0.02***	-0.04***
	(-0.67)	(-0.63)	(-0.64)	(-3.23)	(-4.05)
ROA	-0.01	-0.01	-0.01	-0.03***	-0.11***
	(-0.39)	(-0.35)	(-0.39)	(-3.85)	(-6.11)
Ln (firm size)	0.003	-0.001	0.001	0.01	0.03***
	(0.38)	(-0.19)	(0.16)	(0.32)	(3.56)
1/Share price	2.97***	2.95***	2.96***	3.25***	3.74***
	(8.11)	(8.05)	(8.06)	(10.31)	(10.54)
Ln (turnover)	0.12***	0.12***	0.13***	0.15***	0.16***
	(7.76)	(7.71)	(7.77)	(17.47)	(8.3)
Log(return volatility)	2.48***	2.47***	2.47***	1.93***	1.83***
	(85.28)	(85.24)	(85.16)	(89.67)	(43.63)
NASDAQ	-0.03	-0.02	-0.02	0.01	0.01
	(-1.24)	(-1.11)	(-1.11)	(0.94)	(0.2)
Year	Y	Y	Y	Y	Y
Industry	Y	Y	Y	Y	Y
Adjust cluster errors	N	N	N	Y	Y
Exclude missing advertising	N	N	N	N	Y
Observation	14,690	14,690	14,690	14,690	6,742
Adj. R^2	0.794	0.794	0.794	0.56	0.55

Panel C. Active attention and Amihud Illiquidity Measures

This panel presents the results based on the Amihud (2002) liquidity measures. *ILLIQ* is the monthly average for the daily ratio of absolute return to the dollar volume of the stock. *ILLIQMA* is the ratio of the variable *ILLIQ* to its monthly mean across all stocks. *SD(ILLIQ)* is the monthly standard deviation of *ILLIQ* (Lang and Maffett, 2011). We correct for standard errors using clustering.

	Dependent Variable		
	ILLIQ	*ILLIQMA*	*SD(ILLIQ)*
	Model I	Model II	Model III
Change in SVI	0.02	0.01	0.01
	(1.01)	(1.4)	(0.22)
Ln (number of News)	0	0	0
	(0.81)	(0.81)	(0.79)
Ln (advertising)	0	0	0
	(1.37)	(1.38)	(1.36)
Ln (firm age)	-0.01	-0.01	-0.01
	(-1.16)	(-1.19)	(-1.12)
ROA	-0.02	-0.02	-0.02
	(-1.33)	(-1.32)	(-1.34)
Ln (Firm size)	-0.02	-0.02	-0.02
	(-1.48)	(-1.51)	(-1.47)
1/Share price	-0.08	-0.1	-0.1
	(-1.17)	(-1.17)	(-1.24)
Ln (turnover)	-0.14	-0.16	-0.16
	(-1.52)	(-1.51)	(-1.52)
Ln (return volatility)	0.1	0.11	0.11
	(1.53)	(1.47)	(1.53)
NASDAQ	0.04	0.04	0.04
	(1.44)	(1.43)	(1.53)
Year	Y	Y	Y
Industry	Y	Y	Y
Observation	14690	14690	14690
Adj R^2	0.2423	0.1561	0.2311

Table 5. Robustness Checks

This table reports the estimates from panel regressions relating the alternative liquidity measures to active investor attention (SVI). The sample includes the constituents of the S&P 500 over a period of six years from January 2004 to December 2009. The effective spread is defined as twice the difference between the transaction price and the spread midpoint. The relative effective spread is defined as the effective spread divided by the midpoint of the spread. Share turnover is constructed from CRSP as the monthly average of the share volume divided by the shares outstanding. The SVI is the search frequency of a stock ticker, obtained from Google Trends. The change in SVI is the difference between the SVI in week *t* and the median value of the SVI during the previous eight weeks. The "number of news" is the online media coverage from Google News. Advertising expenditure is obtained from Compustat. The relative bid-ask spread is the monthly average of the ratio of the daily inside spread to the midpoint of the daily inside spread, obtained from CRSP. Firm age is the number of years for which the firm has been included in CRSP. Return on assets is constructed from Compustat as the annual operating income before depreciation, scaled by total assets. Firm size is the market capitalization, calculated as the product of the total number of outstanding shares and the annual closing price. Return volatility is the monthly average of the standard deviation of daily returns, obtained from CRSP. NASDAQ equals 1 for firms listed on the NASDAQ, and 0 otherwise. *, ** and *** represent significance at the 10%, 5% and 1% levels. Standard errors are adjusted for heteroscedasticity.

Panel A. SVI Change and Relative Spread

	Dependent Variables	
	Effective spread	Relative effective spread
	Model I	Model II
Change in SVI	-0.11**	-0.01***
	(-2.02)	(-2.47)
Ln (number of News)	-0.01***	-0.001***
	(-3.07)	(-3.87)
Ln (advertising)	-0.02***	-0.001***
	(-4.97)	(-5.46)
Ln (firm age)	0.05***	0.001***
	(3.87)	(4.74)
ROA	0.11***	0.01***
	(4.75)	(5.22)
Ln (Firm size)	0.13***	0.01***
	(5.28)	(5.67)
1/Share price	0.62***	0.03***
	(3.19)	(3.91)
Ln (turnover)	1.00***	0.05***
	(5.32)	(5.8)
Ln (return volatility)	-0.68***	-0.03***
	(-5.39)	(-5.79)
NASDAQ	-0.25***	-0.01***
	(-5.21)	(-5.66)
Year	Y	Y
Industry	Y	Y
Observation	14,690	14,690
Adj. R^2	0.79	0.79

Panel B. SVI Change and Turnover

	Dependent Variable: Turnover		
	Model I	Model II	Model III
Change in SVI	0.12***	0.12***	0.12***
	(3.8)	(3.8)	(3.8)
Ln (number of News)		0.01***	0.01***
		(3.59)	(2.99)
Ln (advertising)			0.01***
			(10.44)
Ln (firm age)	-0.06***	-0.05***	-0.05***
	(-9.47)	(-9.41)	(-9.37)
ROA	-0.06***	-0.06***	-0.06***
	(-6.71)	(-6.60)	(-6.31)
Ln (Firm size)	-0.08***	-0.08***	-0.09***
	(-16.87)	(-15.79)	(-17.37)
1/ Share price	-0.65***	-0.68***	-0.71***
	(-5.87)	(-6.00)	(-6.21)
Ln (return Volatility)	0.65***	0.65***	0.65***
	(62.04)	(62.01)	(62.25)
NASDAQ	0.27***	0.28***	0.27***
	(20.62)	(21.13)	(21.03)
Year	Y	Y	Y
Industry	Y	Y	Y
Observation	14,690	14,690	14,690
Adj. R^2	0.483	0.484	0.487

Panel C. Substantial Increase in SVI and Propensity Score Matching

This panel presents the results based on the propensity score matching. Substantial increase in SVI (*SIn_SVI*) is set to 1 if the change in SVI is above the top 10 percentile and 0 otherwise. Matching sample of 3002 observation is constructed by Model 1, with 1501 firms as treatment group (*SIn_SVI* =1) and 1501 firms as control group (*SIn_SVI* =0).

	Substantial increase in SVI	Dependent Variable: Ln (Number of Shareholders in thousand)	Relative spread in %
	Model I	Model II	Model III
SIn_SVI		0.14***	-0.08**
		(2.59)	(-2.52)
Ln (number of News)	-0.02***	0.04***	0
	(-2.79)	(2.64)	(-0.25)
Ln (advertising)	0.01*	0.02***	-0.01*
	(1.87)	(2.76)	(-1.92)
Ln (firm age)	0.03*	0.36***	-0.03*
	(1.7)	(7.51)	(-1.66)
ROA	0.06**	0.60***	-0.01
	(2.06)	(9.71)	(-0.16)
Ln (Firm size)	0.11***	0.92***	-0.03
	(7.13)	(31.16)	(-1.54)
1/Share price	-0.87**	15.30***	5.09***
	(-2.01)	(10.70)	(5.96)
Ln (turnover)	0.06**	-0.12	0.08**
	(2.05)	(-1.56)	(2.2)
Ln (return volatility)	0.15***	-0.36***	2.47***
	(4.21)	(-4.53)	(38.33)
NASDAQ	0.30***	-0.71***	-0.13***
	(7.63)	(-7.96)	(-2.65)
Year	Y	Y	Y
Industry	Y	Y	Y
Observation	14690	3002	3002
Adj R^2	0.0214	0.4739	0.7873

Panel D. Corporate Events and Transparency

This panel presents the results after controlling for corporate events or financial reporting transparency. *Event_split* that equals 1 if share splits happened in the month and 0 otherwise; similarly, *Event_dividend* equals 1 if the dividend distribution was announced in the month and 0 otherwise.

	Dependent Variable				
	Ln(Number of Shareholders in thousand)		Relative spread in %		
	Model I	Model II	Model III	Model IV	Model V
Change in SVI	0.21*	0.27**	-0.24***	-0.23***	-0.23***
	(1.7)	(2.07)	(-3.43)	(-3.01)	(-3.28)
Ln (number of News)	0.04	0.03	0.01	0.01	0
	(0.88)	(0.78)	(1.02)	(1.02)	(0.27)
Ln (advertising)	0.01	0	0	0	0
	(0.24)	(0.04)	(-0.74)	(-0.94)	(-0.65)
Ln (firm age)	0.27**	0.25*	-0.01	0	0
	(2.01)	(1.96)	(-0.44)	(-0.21)	(-0.04)
Return	-0.20*	-0.27**			
	(-1.75)	(-2.31)			
ROA	0.31	0.35*	-0.01	0	0
	(1.52)	(1.75)	(-0.21)	(-0.12)	(0.01)
Ln (Firm size)	0.87***	0.88***	0	0	0.01
	(10.4)	(10.28)	(0.04)	(-0.04)	(0.45)
1/Share price	8.13***	8.28***	2.97***	2.87***	2.92***
	(3.35)	(3.42)	(4.76)	(4.62)	(4.6)
Ln (turnover)	-0.08	-0.11	0.12***	0.13***	0.14***
	(-0.47)	(-0.64)	(2.78)	(2.77)	(3.2)
Ln (return volatility)	-0.13	-0.15	2.48***	2.50***	2.44***
	(-1.09)	(-1.11)	(36.71)	(35.36)	(34.07)
NASDAQ	-0.64**	-0.64**	-0.02	-0.02	-0.02
	(-2.11)	(-2.05)	(-0.55)	(-0.54)	(-0.4)
Event_dividend	0.08		0.04*		
	(0.77)		(1.89)		
Event_split	-0.17		-0.14		
	(-0.59)		(-1.4)		
Big4					-0.24
					(-1.3)
Year	Y	Y	Y	Y	Y
Industry	Y	Y	Y	Y	Y
Adjust cluster errors	Y	Y	Y	Y	Y
Exclude event months	N	Y	N	Y	N
Obs	14690	12591	14690	12591	14690
Adj. R^2	0.3981	0.3906	0.7937	0.7918	0.7923

Appendix

Panel A. Active Attention and Breadth of Ownership: Univariate Analysis

This panel reports the number of shareholders (in thousand) for stocks with low and high attention. The sample includes the constituents of S&P 500 over a period of 6 years from January 2004 to December 2009. The attention is measured by SVI, the search frequency of stock ticker from Google Trends. The firms are further classified into small and large ones based on the mean of market capitalization. *, ** and *** represent significance at the 10%, 5% and 1% level.

	SVI		Difference	*t-statistics*
	Low	High		
All	46.96	94.27	47.31***	14.38
Small	22.45	24.72	2.27**	2.06
Large	78.86	153.24	74.38***	11.54

Panel B. Active Attention and Stock Liquidity: Univariate Analysis

This panel reports presets the relative bid-ask spread (in %) for stocks with low and high attention. The sample includes the constituents of S&P 500 over a period of 6 years from January 2004 to December 2009. The attention is measured by SVI, the search frequency of stock ticker from Google Trends. The firms are further classified into small and large ones based on the mean of market capitalization. *, ** and *** represent significance at the 10%, 5% and 1% level.

	SVI		Difference	t-statistics
	Low	high		
All stock	3.08	2.89	-0.19***	6.38
Small	3.44	3.27	-0.17***	3.43
Large	2.62	2.56	-0.06*	1.72

Panel C. Firm Fixed Effects

This panel reports the estimates from panel regressions with firm fixed effects applied.

	Dependent Variable			
	Ln(Number of Shareholders in thousand)		Relative spread in %	
	Model I	Model II	Model III	Model IV
Change in SVI	0.01**	0.02	-0.03***	-0.24***
	(2.11)	(0.64)	(-3.57)	(-4.44)
Ln (number of News)	0.01	0.01	0.04**	0.04**
	(0.83)	(0.74)	(2.36)	(2.53)
Ln (advertising)	-0.06***	-0.06***	-0.01	-0.01
	(-12.28)	(-12.2)	(-0.96)	(-1.13)
Ln (firm age)	-0.01	-0.01	0.27***	0.27***
	(-0.22)	(-0.27)	(4.17)	(4.24)
Return	0.05	0.05		
	(1.49)	(1.47)		
ROA	0.08***	0.08***	0.03	0.03
	(4.45)	(4.4)	(0.74)	(0.8)
Ln (Firm size)	0.02	0.02	-0.14***	-0.14***
	(1.15)	(1.2)	(-4.88)	(-4.89)
1/Share price	0.95***	0.94***	2.08***	2.11***
	(6.57)	(6.53)	(7.53)	(7.65)
Ln (turnover)	0.10***	0.10***	0.14***	0.14***
	(7.73)	(7.75)	(5.74)	(5.87)
Ln (return volatility)	-0.04***	-0.04***	2.48***	2.48***
	(-3.91)	(-3.9)	(121.89)	(121.97)
NASDAQ	omitted	omitted	omitted	omitted
Year	Y	Y	Y	Y
Industry	Y	Y	Y	Y
Firm fixed effects	Y	Y	Y	Y
Obs.	14690	14690	14690	14690
Overall R^2	0.0198	0.0209	0.7674	0.7673